

AY ÖÜÖÄ WT ÓÜÜ: W81XWH-12-1-0115

TITLE:

Systematic Characterization of the Molecular Mechanisms That Regulate and Mediate
Alternative Lengthening of Telomeres in Breast Carcinoma

PRINCIPAL INVESTIGATOR: Yaara Zwang

CONTRACTING ORGANIZATION:

Dana-Farber Cancer Institute

Boston, MA 02115-6013

REPORT DATE: April 2014

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command

Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;

Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE April 2014			2. REPORT TYPE Annual Summary		3. DATES COVERED 15/March/2013-14/March/2014	
4. TITLE AND SUBTITLE: Systematic Characterization of the Molecular Mechanisms That Regulate and Mediate Alternative Lengthening of Telomeres in Breast Carcinoma					5a. CONTRACT NUMBER W81XWH-12-1-0115	
					5b. GRANT NUMBER W81XWH-12-1-0115	
					5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Yaara Zwang E-Mail: yaarac_zwang@dfci.harvard.edu					5d. PROJECT NUMBER	
					5e. TASK NUMBER	
					5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Dana-Farber Cancer Institute 450 Brookline Ave Boston, MA 02115-6013					8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012					10. SPONSOR/MONITOR'S ACRONYM(S)	
					11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited						
13. SUPPLEMENTARY NOTES						
14. ABSTRACT Alternative lengthening of telomeres (ALT) is a mechanism utilized by several cancer types to maintain telomeres. The aim of this project is to characterize the mechanisms that mediate and regulate ALT activity in breast cancer, by utilizing proteomic and genomic approaches to describe and perturb ALT+ cells. Gaining this knowledge on ALT activity can promote the development of novel treatment for ALT+ breast tumors, and patients with telomerase inhibitor resistant tumors, while having minimal non-desired effects on the normal cell population that typically has no ALT activity. Here, I report on my progress in profiling the global transcriptional patterns that differentially characterize ALT+ cells. I describe the analytical tools that I used to define an ALT-related gene signature, and my efforts to identify potential functional relationship to ALT activity.						
15. SUBJECT TERMS none provided						
16. SECURITY CLASSIFICATION OF:				17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	19b. TELEPHONE NUMBER (include area code)			
				UU	20	

Table of Contents

Introduction	4
Body	5
Unbiased analysis of transcriptional profiles of ALT+ and non-ALT cells.....	5
Studying the differential expression pattern of ALT+ cells	5
Key research accomplishments	8
Conclusion.....	9
References	10
Supporting data	13
Legend to figures	20

Introduction

The capacity to maintain elongated telomeres is an essential feature of tumor cells enabling their unlimited replicative potential. Approximately 85-90% of human cancers maintain telomeres through activation of telomerase. However, several cancer types harness telomerase-independent mechanisms referred to as Alternative Lengthening of Telomeres (ALT). Importantly, inhibition of telomerase in immortalized premalignant cells and in cancer cells induces a switch from telomerase activation to ALT, enabling the cells to escape therapy (Queisser et al, 2013; Villa et al, 2000).

The phenotypic characteristics of ALT+ cells include long and heterogeneous telomeric DNA, ALT-associated PML Bodies (APBs), high level of recombination at telomeres, and abundant extrachromosomal telomeric DNA, especially single strand C-rich circles (C-circles)(Cesare & Reddel, 2010). Currently, there is limited molecular understanding of the mechanisms that induce ALT and that are involved in ALT activity.

Importantly, ALT activity has been reported in breast carcinomas, where it was correlated with HER2 positivity and with poor prognosis (Subhawong et al, 2009). Hence, understanding the mechanisms that regulate and facilitate ALT activity can promote the development of novel treatments for telomerase-independent breast cancer. Such treatment has an attractive potential in helping patients with ALT+-poor prognosis breast cancer and as a combination therapy for patients treated with telomerase inhibitors. It is expected that inhibition of ALT activity will have minimal non-desired effects on the normal cell population that typically has no ALT activity.

The aim of this project is to characterize the mechanisms that mediate and regulate ALT activity in breast cancer, by utilizing proteomic and genomic approaches to describe and perturb ALT+ cells.

Body

Unbiased analysis of transcriptional profiles of ALT+ and non-ALT cells

In order to characterize the unique transcriptional profile of ALT+ cells, I measured the steady-state global transcriptional program of 16 ALT+ and 9 non-ALT cell-lines using the Affymetrix microarray HTU133A platform.

Data was extracted from raw files using the RMA method, profiling 22,277 probes across all 25 cell-lines. To allow comparison of expression between cell-lines, the data was normalized (Z-score normalization) across all samples using a GenePattern module ('PreprocessDataset') (Reich et al, 2006). In order to visualize the variance in the data, I used Principal Component Analysis (PCA), which uses multivariate analysis to reveal the internal structure of the data. The main advantage of this analysis is that it highlights global inter-variable correlations in an unbiased manner, thus allowing the identification of major sub-groups of cell-lines within the complete data set that exhibit higher correlation to each other. As shown in Figure 1A, the first two PCAs with the highest variance clustered the cell-lines into two major groups according to their ALT activity (variation across PCA#1 is 16.56%; variation across PCA#2 is 9.16%). Notably, 13/16 ALT+ cell-lines (green) exhibited relatively higher values along PCA#1 and lower values along PCA#2 (marked by ellipse I). 8/9 non-ALT cell-lines (orange) exhibited lower values along PCA#1 and higher values along PCA#2 (marked by ellipse II). Importantly, this analysis shows that the ALT+ cell-lines correlate better with each other than with non-ALT cell-lines. Furthermore, this analysis also indicates that the most significant feature that distinguishes the samples is their ALT activity, since the two most variant PCAs were able to distinguish the ALT+ from the non-ALT cell-lines.

To explore the unique transcriptional profile related to ALT activity, I first identified the genes that exhibit variability within the data using GenePattern. A total of 9,909 probes exhibited over two-fold change and a minimal delta of 50 across cell-lines (pre-normalized data), independent of ALT-activity status. In accordance with the PCA analysis (Figure 1A), a correlation matrix of all the samples showed two major clusters distinguishing cell-lines by their ALT activity (Figure 1B). The correlation coefficients presented here are affected by cell-line specific expression patterns as well as by ALT-related patterns since all the genes that exhibited variability across all samples were included in computing this correlation matrix. This suggests that the contribution of ALT-related expression patterns is significant enough to lead an unbiased clustering of the measured cell lines.

Studying the differential expression pattern of ALT+ cells

To facilitate better understanding of the molecular mechanisms that are related to ALT activity, I used the GenePattern Comparative Marker Selection module to identify genes that exhibit differential expression between ALT+ and non-ALT cell-lines (Gould et al, 2006). This analysis computes a score for each gene that corresponds to the correlation between the gene's expression and the phenotypic classification of the cell-lines by their ALT activity. The genes

were then ranked-ordered by this score (Figure 2A). To correct for multiple hypothesis testing, the Comparative Marker Selection tool calculates a false-discovery rate value for each gene based on permutations of the cell-lines phenotypic classification (Figure 2B). Based on this analysis, I identified 5022 genes whose differential expression significantly ($q\text{-value} < 0.01$) correlates with ALT activity, of which 2147 genes had a $q\text{-value} < 0.001$ (corresponding to $\text{FDR} < 0.1\%$; Figure 2C).

Investigating the list of differentially expressed genes revealed that several known ALT-related genes were indeed up-regulated in ALT+ cells. Members of the MRN complex, which is essential for telomere elongation in ALT+ cell-lines (Jiang et al, 2005), MRE11, RAD50, and XRCC1, were significantly up-regulated in ALT+ compared to non-ALT cell-lines (Figure 2D,E,M). In addition, SMC5, SMC6, MUS81, Sp100, FANCL and FEN1 were also significantly up-regulated in the ALT+ cell-lines (Figure 2F-J,P), in accordance to their reported contribution to ALT activity (Potts & Yu, 2007; Saharia & Stewart, 2009; Zeng et al, 2009). The telomeric heterochromatin proteins, TREF1 and TREF2 were also significantly up-regulated in ALT+ cells (Figure 2N-O). Additionally, a recent report suggested that ASF1B, an histone chaperone, can induce ALT phenotype in both primary as well as cancer cell (O'Sullivan et al, 2014). In accordance, the transcription of ASF1B was significantly up-regulated in the ALT+ cell-lines (Figure 2K). Similarly, RECQL, a member of the BLM/WRN complex that was recently shown to be essential for telomere elongation in ALT-active cells (Popuri et al, 2014), was also up-regulated in ALT+ cell-lines (Figure 2L).

To further study the unique gene expression pattern of ALT+ cell-lines, I took an unbiased approach to analyze the gene signature identified above by using GSEA (Gene Set Enrichment Analysis) tool. First, I analyzed the enrichment of the ALT-related signature genes against all Gene Ontology (GO) gene sets. Each of the sets in this collection consists of genes that are annotated by the same GO term. 84 out of the 1025 gene sets included in this collection were significantly enriched for in the ALT+ cell-lines, of which 24 gene sets had a nominal $p\text{-value}$ lower than 1%. Several of the enriched gene sets were associated with DNA replication, mitosis, chromosome segregation, and spindle regulation (Table 1). This enrichment of cell-cycle-related genes could be a general feature of immortalized cells. However, in this data set, the observed enrichment associates with ALT-related immortalization, while non-ALT, hTERT-immortalized cell-lines exhibit significantly lower expression of these genes (see a representative gene set in Figure 3A). Interestingly, one of the top enriched gene sets identified here is termed 'ATP DEPENDENT DNA HELICASE ACTIVITY' (Figure 3B). This set includes genes whose activity relates to the MRN complex as well as to BLM and WRN, suggesting a role in ALT activity.

A second gene set collection I used consists of curated gene sets that are based on data from the literature, pathway databases (such as KEGG and REACTOM), and perturbation datasets. 307 out of the 3767 gene sets included in this collection were significantly enriched for in the ALT+ cell-lines, of which 153 gene sets had a nominal $p\text{-value}$ lower than 1% (Table 2). Several of the enriched gene sets identified in this analysis are related to DNA synthesis and

more specifically, to the synthesis of the lagging strand (Figure 3C). The disabled synthesis of the lagging strand of telomeres in normal cells leads to the end-replication problem that ultimately results in shortening of the telomeres. In the absence of telomerase, ALT+ cells must activate mechanisms to elongate and maintain telomeres. Thus, higher expression of genes associated with the synthesis of the DNA lagging strand could contribute to telomere maintenance. Confirming the relevance of this gene set to ALT activity, the set includes FEN1, which is essential for stabilizing telomeres in ALT+ cells but not in normal or telomerase-active cells (Saharia & Stewart, 2009). A subset of this gene set, consisting of genes that are specifically involved in the removal of the flap intermediate during Okazaki fragments joining, was also significantly enriched for in the ALT+ cell-lines (Figure 3D). Additional enriched gene sets that could potentially be relevant to ALT+ cell-lines are sets related to DNA repair and maintenance of stability (Figure 3E). ALT activity involves extensive recombination at the chromosomal ends, and thus could increase the overall burden of DNA repair in the cell. Therefore, over-expression of repair genes could be essential for maintaining DNA stability in ALT+ cells. Of note, several of the genes included in the set, such as AURKA, are already being tested as targets for targeted therapy in breast cancer, suggesting ALT+ breast tumors could benefit from these drugs (Romanelli et al, 2012; Sun et al, 2008).

Finally, a gene set consisting of genes encoding kinases that are differentially expressed in basal breast cancer was enriched for in the ALT+ cell-lines (Figure 3F) (Finetti et al, 2008). Most of these kinases are associated with cell-cycle regulation, such as AURKA, AURKB, BUB1, and CDK1. Interestingly, the expression of PLK1 was reported to increase as breast cells undergo immortalization (Li et al, 2007). PLK1 was also reported to contribute to ALT activity by phosphorylating TREF1, thus promoting its binding and stabilization of telomeres (Wu et al, 2008). Likewise, CDK1 regulates the replication of telomeric DNA in telomerase-negative cells (Dai et al, 2012), and also phosphorylates TREF1 to promote its recruitment to sites of DNA damage (McKerlie et al, 2013).

The differential expression of these genes between ALT+ and non-ALT cell-lines could potentially be utilized to gain therapeutic specificity, however, further investigation of the contribution of these kinases to ALT activity is necessary.

To summarize, ALT+ cell-lines exhibit a differential expression pattern as compared to non-ALT cell-lines. This ALT-related expression signature includes several known ALT-related genes, which signifies its relevance to understanding the mechanisms that mediate and regulate ALT activity. Using statistical analysis tools, I identified several sets of genes that are differentially expressed in ALT+ cell-lines and are potentially involved in ALT activity and the survival of ALT+ cells.

Key research accomplishments

- Profiled gene expression patterns of 16 ALT+ cell-lines as well as 9 non-ALT cell-lines.
- Defined an ALT-related gene expression signature.
- Identified several sets of genes that are uniquely expressed in ALT+ cells and are potentially related to ALT activity.

Conclusion

Profiling global gene expression from 16 ALT+ cell-lines and 9 non-ALT cell-lines revealed an ALT-related differential gene expression signature. Importantly, I show here that these profiles dictate clustering of the cell-lines according to their ALT activity in an unbiased manner, suggesting that ALT activity imposes a significant and wide effect on global gene expression patterns. I could confirm the relevance of the expression signature to ALT activity by showing that multiple genes that are known to be involved in ALT are also differentially up-regulated in ALT+ cell-lines.

Mining the ALT-related expression signature using Gene Set Enrichment Analysis identified several sets of genes with potential relevance to ALT activity. Interestingly, many of these genes regulate DNA replication, the cell-cycle, and DNA damage repair. The up-regulation of those genes in ALT+ cell-lines suggests that ALT activity involves re-wiring of these exiting mechanisms to elongate and maintain telomeres in the absence of telomerase. It could also indicate potential dependencies of ALT+ cells, such as increased DNA damage repair activity, in line with our previous observation of high level of genomic instability and frequent presence of DNA damage in ALT+ cells (Lovejoy et al, 2012).

This comprehensive profiling of the ALT-related gene expression is a useful resource for studying genes that are involved in ALT activity. Several of the gene sets identified here include known drug targets and thus could have important clinical implications.

References

Cesare AJ, Reddel RR (2010) Alternative lengthening of telomeres: models, mechanisms and implications. *Nature reviews Genetics* **11**: 319-330

Dai X, Huang C, Chai W (2012) CDK1 differentially regulates G-overhang generation at leading- and lagging-strand telomeres in telomerase-negative cells in G2 phase. *Cell cycle* **11**: 3079-3086

Finetti P, Cervera N, Charafe-Jauffret E, Chabannon C, Charpin C, Chaffanet M, Jacquemier J, Viens P, Birnbaum D, Bertucci F (2008) Sixteen-kinase gene expression identifies luminal breast cancers with poor prognosis. *Cancer research* **68**: 767-776

Gould J, Getz G, Monti S, Reich M, Mesirov JP (2006) Comparative gene marker selection suite. *Bioinformatics* **22**: 1924-1925

Jiang WQ, Zhong ZH, Henson JD, Neumann AA, Chang AC, Reddel RR (2005) Suppression of alternative lengthening of telomeres by Sp100-mediated sequestration of the MRE11/RAD50/NBS1 complex. *Molecular and cellular biology* **25**: 2708-2721

Li Y, Pan J, Li JL, Lee JH, Tunkey C, Saraf K, Garbe JC, Whitley MZ, Jelinsky SA, Stampfer MR, Haney SA (2007) Transcriptional changes associated with breast cancer occur as normal human mammary epithelial cells overcome senescence barriers and become immortalized. *Molecular cancer* **6**: 7

Lovejoy CA, Li W, Reisenweber S, Thongthip S, Bruno J, de Lange T, De S, Petrini JH, Sung PA, Jasin M, Rosenbluh J, Zwang Y, Weir BA, Hatton C, Ivanova E, Macconail L, Hanna M, Hahn WC, Lue NF, Reddel RR, Jiao Y, Kinzler K, Vogelstein B, Papadopoulos N, Meeker AK, Consortium ALTSC (2012) Loss of ATRX, genome instability, and an altered DNA damage response are hallmarks of the alternative lengthening of telomeres pathway. *PLoS genetics* **8**: e1002772

McKerlie M, Walker JR, Mitchell TR, Wilson FR, Zhu XD (2013) Phosphorylated (pT371)TRF1 is recruited to sites of DNA damage to facilitate homologous recombination and checkpoint activation. *Nucleic acids research* **41**: 10268-10282

O'Sullivan RJ, Arnoult N, Lackner DH, Oganessian L, Haggblom C, Corpet A, Almouzni G, Karlseder J (2014) Rapid induction of alternative lengthening of telomeres by depletion of the histone chaperone ASF1. *Nature structural & molecular biology* **21**: 167-174

Popuri V, Hsu J, Khadka P, Horvath K, Liu Y, Croteau DL, Bohr VA (2014) Human RECQL1 participates in telomere maintenance. *Nucleic acids research*

Potts PR, Yu H (2007) The SMC5/6 complex maintains telomere length in ALT cancer cells through SUMOylation of telomere-binding proteins. *Nature structural & molecular biology* **14**: 581-590

Queisser A, Heeg S, Thaler M, von Werder A, Opitz OG (2013) Inhibition of telomerase induces alternative lengthening of telomeres during human esophageal carcinogenesis. *Cancer genetics* **206**: 374-386

Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP (2006) GenePattern 2.0. *Nature genetics* **38**: 500-501

Romanelli A, Clark A, Assayag F, Chateau-Joubert S, Poupon MF, Servely JL, Fontaine JJ, Liu X, Spooner E, Goodstal S, de Cremoux P, Bieche I, Decaudin D, Marangoni E (2012) Inhibiting aurora kinases reduces tumor growth and suppresses tumor recurrence after chemotherapy in patient-derived triple-negative breast cancer xenografts. *Molecular cancer therapeutics* **11**: 2693-2703

Saharia A, Stewart SA (2009) FEN1 contributes to telomere stability in ALT-positive tumor cells. *Oncogene* **28**: 1162-1167

Subhawong AP, Heaphy CM, Argani P, Konishi Y, Kouprina N, Nassar H, Vang R, Meeker AK (2009) The alternative lengthening of telomeres phenotype in breast carcinoma is associated with HER-2 overexpression. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* **22**: 1423-1431

Sun L, Li D, Dong X, Yu H, Dong JT, Zhang C, Lu X, Zhou J (2008) Small-molecule inhibition of Aurora kinases triggers spindle checkpoint-independent apoptosis in cancer cells. *Biochemical pharmacology* **75**: 1027-1034

Villa R, Folini M, Perego P, Supino R, Setti E, Daidone MG, Zunino F, Zaffaroni N (2000) Telomerase activity and telomere length in human ovarian cancer and melanoma cell lines: correlation with sensitivity to DNA damaging agents. *International journal of oncology* **16**: 995-1002

Wu ZQ, Yang X, Weber G, Liu X (2008) Plk1 phosphorylation of TRF1 is essential for its binding to telomeres. *The Journal of biological chemistry* **283**: 25503-25513

Zeng S, Xiang T, Pandita TK, Gonzalez-Suarez I, Gonzalo S, Harris CC, Yang Q (2009)
Telomere recombination requires the MUS81 endonuclease. *Nature cell biology* **11**: 616-623

Supporting data

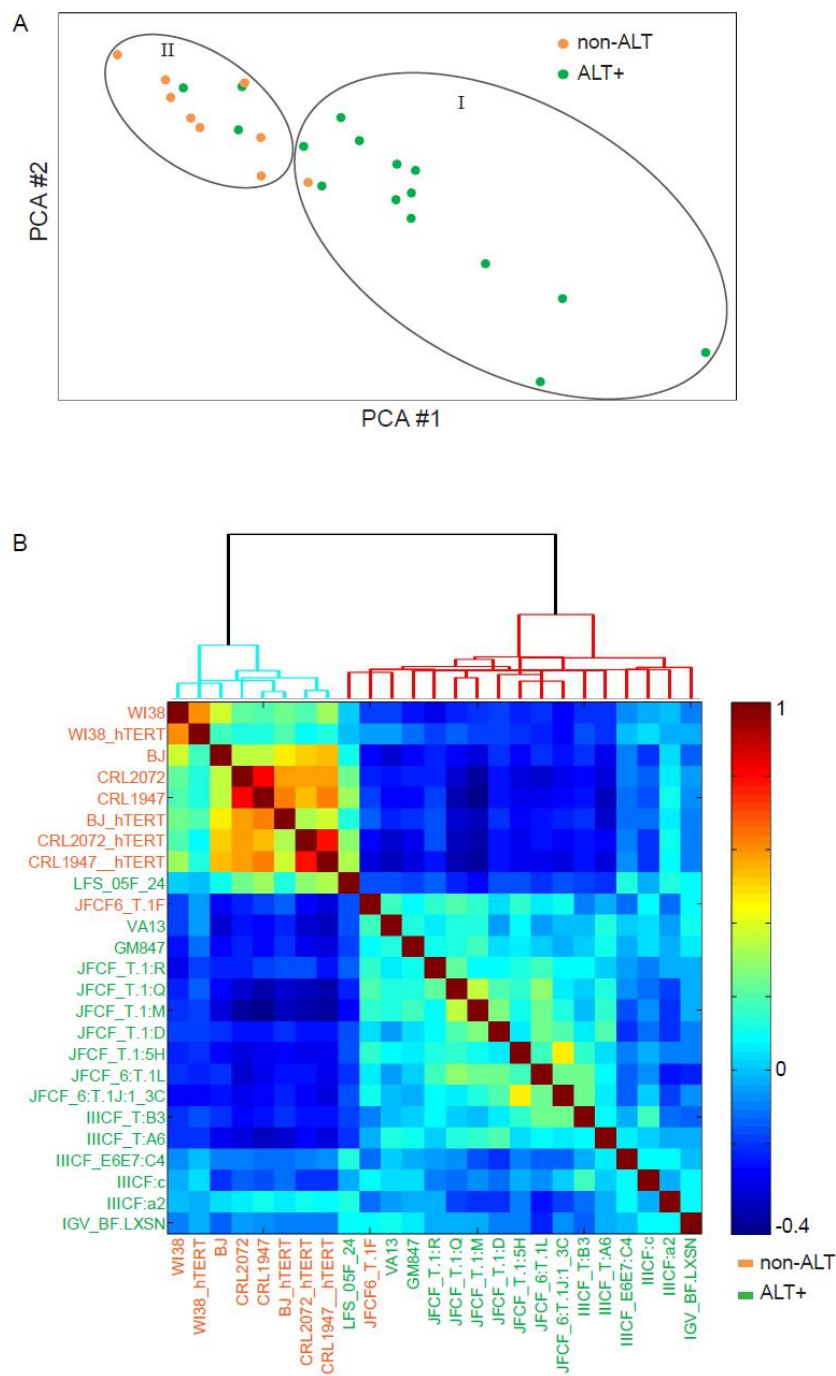


Figure 1

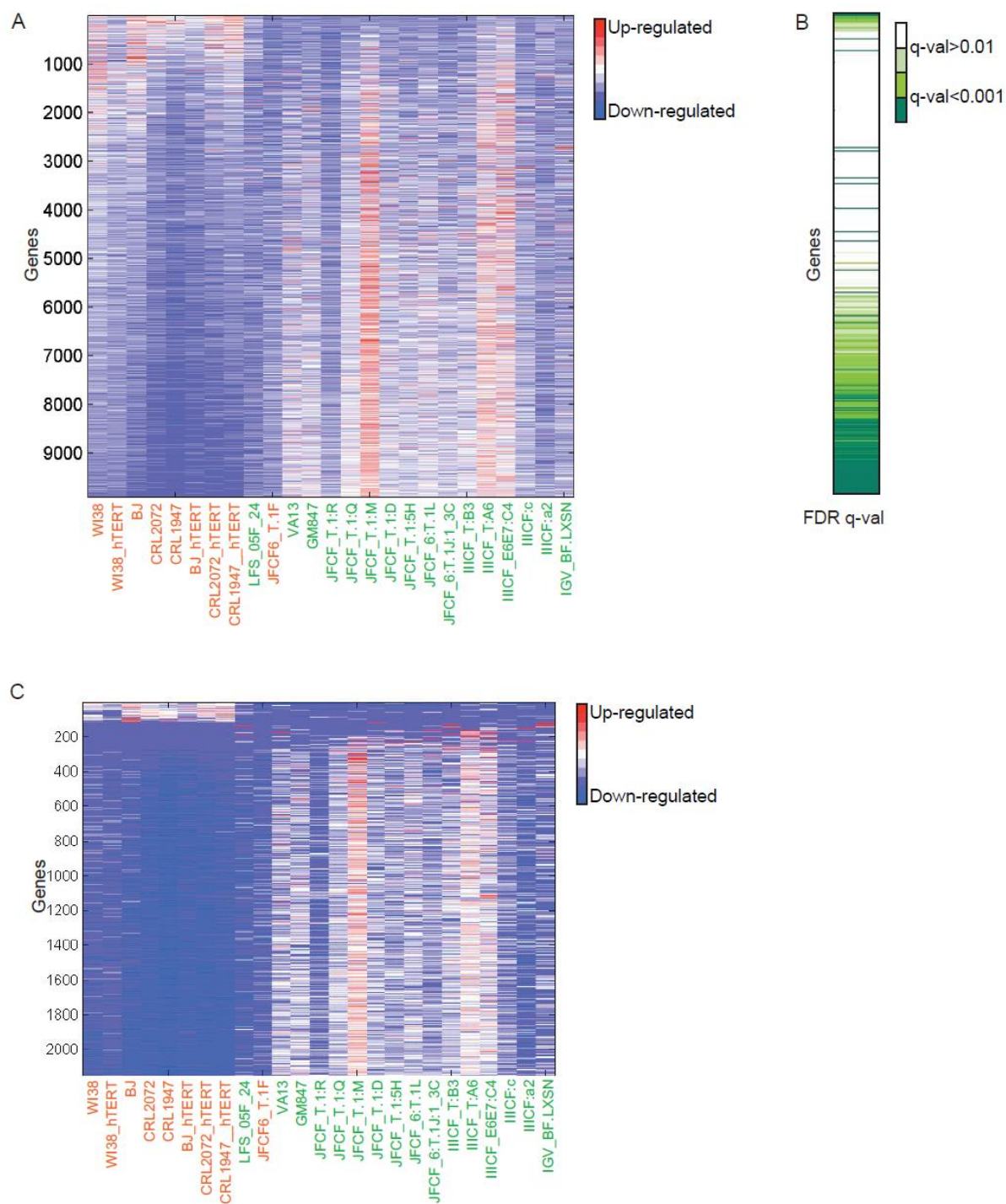


Figure 2

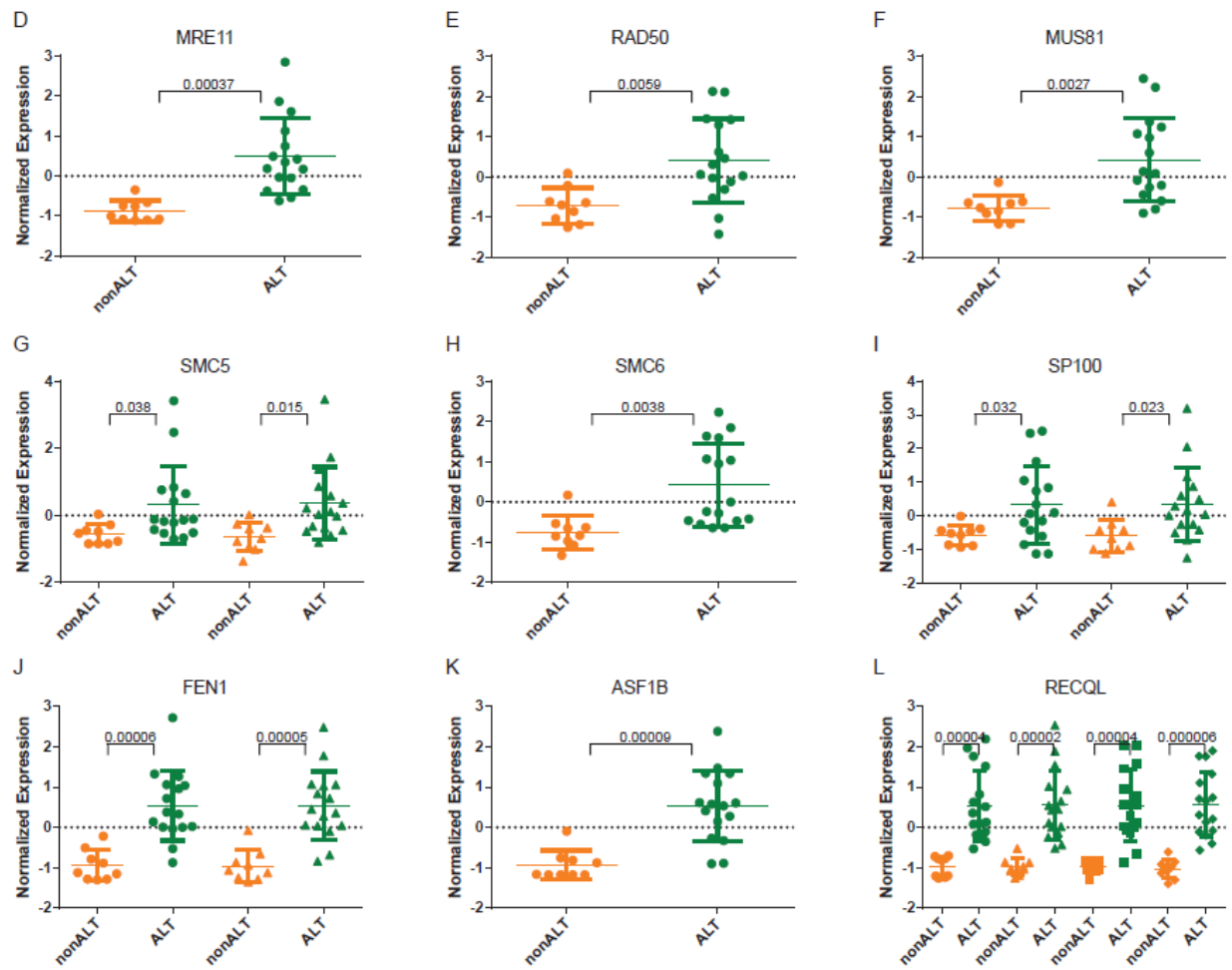


Figure 2 cont.

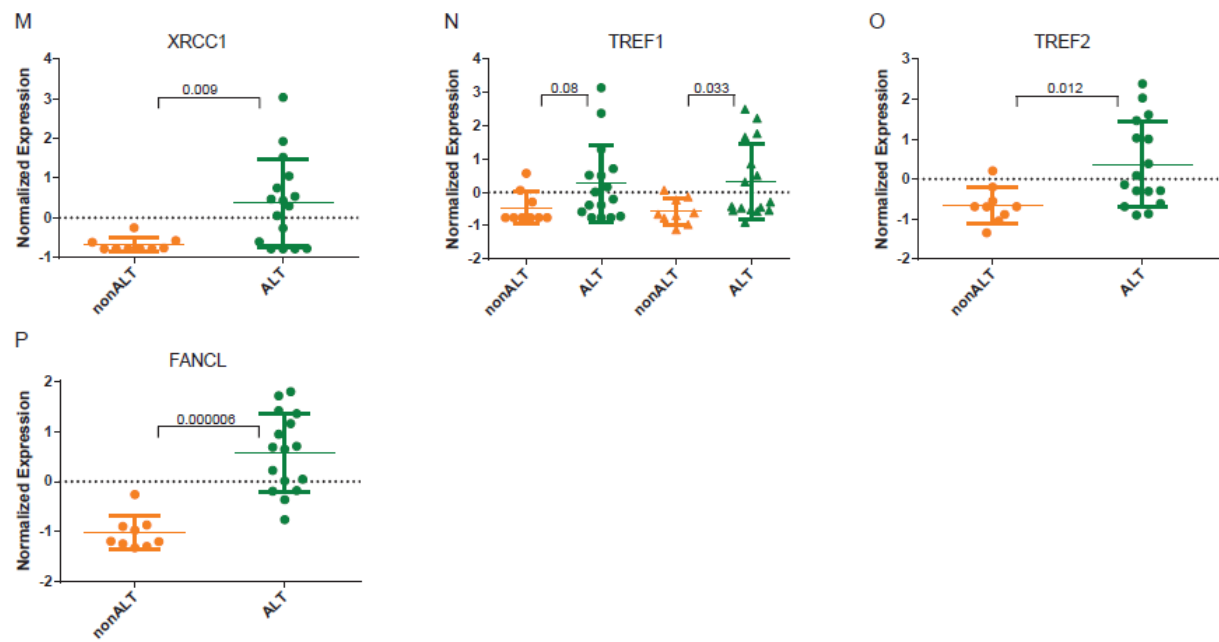


Figure 2 cont.

A. GO: MITOTIC SISTER CHROMATID SEGREGATION

B. GO: ATP DEPENDENT DNA HELICASE ACTIVITY

C. REACTOME: PROCESSIONAL SYNTHESIS ON THE LAGGING STRAND

D. REACTOME: REMOVAL OF THE FLAP INTERMEDIATE FROM THE C STRAND

E. KEGG: MISMATCH REPAIR

F. FINETTI BREAST CANCER KINOME RED

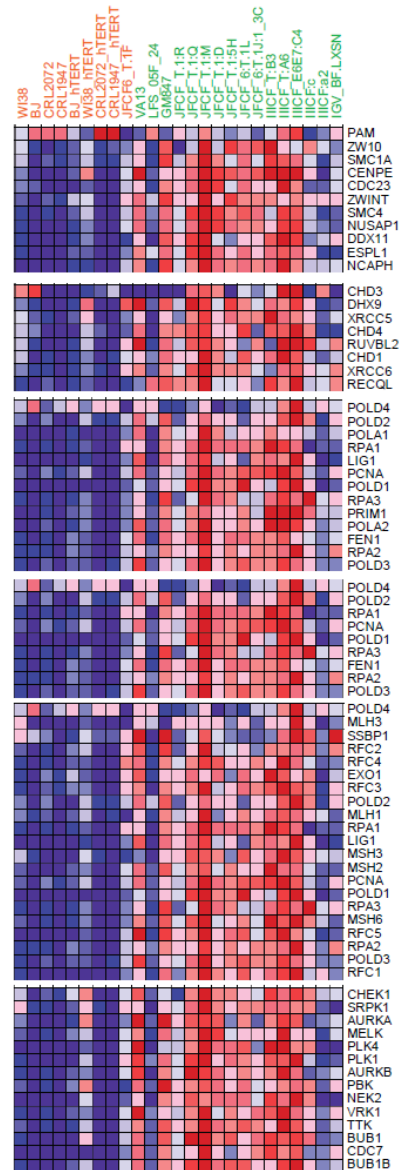


Figure 3

NAME	SIZE	NES
MITOTIC SISTER CHROMATID SEGREGATION	11	-0.77373
SISTER CHROMATID SEGREGATION	11	-0.77373
ATP DEPENDENT DNA HELICASE ACTIVITY	8	-0.74028
G1 PHASE OF MITOTIC CELL CYCLE	8	-0.71781
REPLICATION FORK	14	-0.7006
ORGANELLAR SMALL RIBOSOMAL SUBUNIT	8	-0.68236
MITOCHONDRIAL SMALL RIBOSOMAL SUBUNIT	8	-0.68236
SMALL RIBOSOMAL SUBUNIT	8	-0.68236
CHROMOSOME SEGREGATION	20	-0.67529
REGULATION OF MITOSIS	28	-0.6676
SPINDLE MICROTUBULE	15	-0.66266
MICROTUBULE MOTOR ACTIVITY	12	-0.65116
KINETOCHORE	20	-0.64197
DNA DEPENDENT ATPASE ACTIVITY	17	-0.63868
MITOTIC CELL CYCLE CHECKPOINT	16	-0.63795
G1 PHASE	9	-0.63567
DNA DIRECTED DNA POLYMERASE ACTIVITY	12	-0.62905
MICROTUBULE ORGANIZING CENTER ORGANIZATION AND BIOGENESIS	9	-0.62795
DRUG BINDING	10	-0.62713
DNA POLYMERASE ACTIVITY	14	-0.62648

Table 1: Top 20 GO gene sets enriched for in ALT+ cell-lines. Size indicates the number of genes included in each set. NES, Normalized Enrichment Score. Reflecting the degree to which a gene set is over-represented at either ends of the ranked list, normalized by permutations of the ranked dataset. A more negative NES indicates higher enrichment of the gene set in the ALT+ differentially up-regulated end of the ranked list.

NAME	SIZE	NES
REACTOME PROCESSION SYNTHESIS ON THE LAGGING STRAND	13	-0.89422
REACTOME REMOVAL OF THE FLAP INTERMEDIATE FROM THE C STRAND	9	-0.89365
REACTOME LAGGING STRAND SYNTHESIS	17	-0.79381
REACTOME POL SWITCHING	12	-0.79306
LY AGING MIDDLE DN	16	-0.79292
KANG DOXORUBICIN RESISTANCE UP	47	-0.78178
AMUNDSON GAMMA RADIATION RESPONSE	34	-0.77391
FINETTI BREAST CANCER KINOME RED	14	-0.76779
KALMA E2F1 TARGETS	11	-0.75699
MONTERO THYROID CANCER POOR SURVIVAL UP	10	-0.75409
REACTOME PURINE RIBONUCLEOSIDE MONOPHOSPHATE BIOSYNTHESIS	9	-0.75289
KUMAMOTO RESPONSE TO NUTLIN 3A DN	9	-0.7507
BURTON ADIPOGENESIS PEAK AT 24HR	25	-0.74505
REACTOME REPAIR SYNTHESIS FOR GAP FILLING BY DNA POL IN TC NER	14	-0.74457
KEGG MISMATCH REPAIR	21	-0.744
TANG SENESENCE TP53 TARGETS DN	45	-0.7432
KAMMINGA EZH2 TARGETS	34	-0.73923
CROONQUIST NRAS SIGNALING DN	66	-0.73693
REACTOME SLBP DEPENDENT PROCESSING OF REPLICATION DEPENDENT HISTONE PRE MRNAS	8	-0.73269
CROONQUIST IL6 DEPRIVATION DN	80	-0.7323

Table 2: Top 20 curated gene sets enriched for in ALT+ cell-lines.

Legend to figures

Figure 1: Analysis of gene expression profiles of ALT+ and non-ALT cell-lines shows significant ALT-related difference in global gene expression. **A.** PCA analysis of normalized expression data. The values for the PCAs with the highest variance are plotted for each cell-line. Green, ALT+ cell-lines; Orange, non-ALT cell-lines. The circles labeled I and II mark the two identified clusters. **B.** Correlation coefficients were calculated for all variable genes across all cell-lines and plotted as a heatmap. Red, positive correlation; Blue, Negative correlation. Dendrogram above corresponds to the degree of linkage between cell-lines.

Figure 2: Comparative Marker Selection analysis defines an ALT-related gene signature. **A.** Plotted is the expression of genes across all cell-lines ranked according to the correlation between the gene's expression and the phenotypic classification of the cell-lines. Red, up-regulated; Blue, down-regulated. **B.** The q-values of false-discovery rate for the Comparative Marker Selection analysis are plotted, ranked as in A. **C.** The expression of the 2147 most significantly differentially expressed genes according to the Comparative Marker Selection analysis is plotted. **D-P.** Expression of known ALT-related genes is compared between non-ALT and ALT+ cell-lines. Mean and SD indicated by bar, p-value for differential expression (T-test) is presented. For several genes multiple probes were present in the data, and were plotted accordingly.

Figure 3: GSEA reveals potential associations to ALT activity. A-F. Expression of genes included in the indicated enriched gene sets across all cell-lines. Heatmap colors as in Figure 2A.